

Memory Formation: From Condensed Matter to Biological Matter and Beyond
KITP 2018

**MECHANICS AND INFORMATION
IN A MINIMAL MODEL OF PROTEIN EVOLUTION**

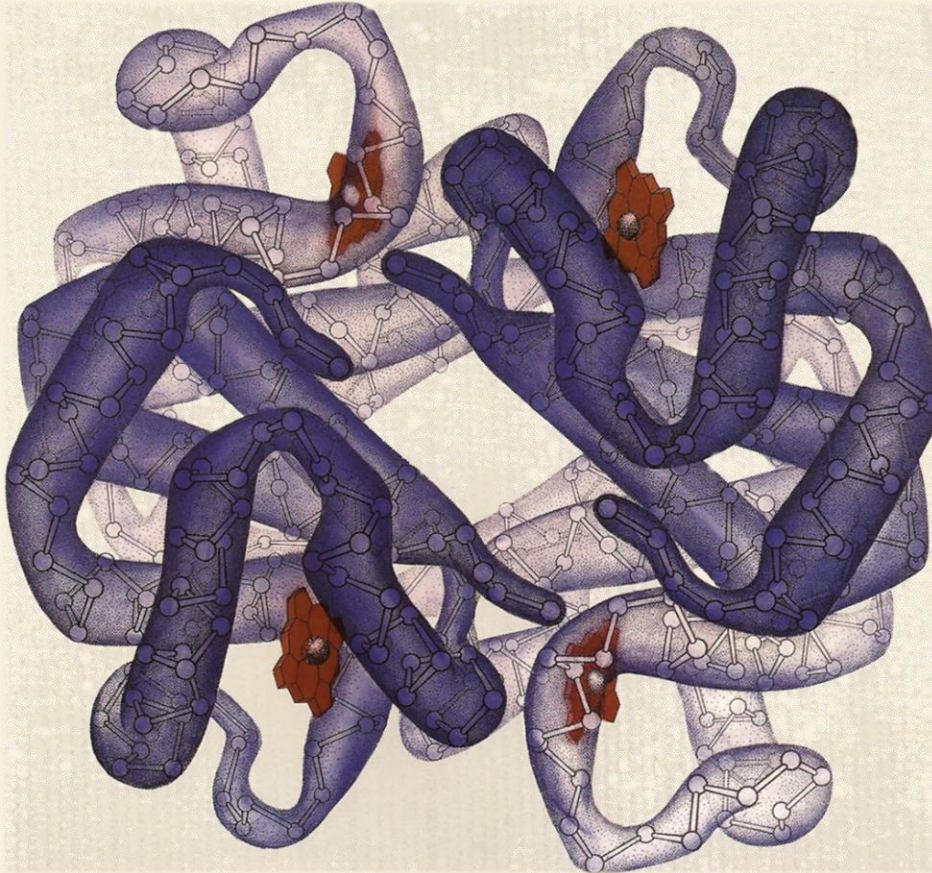
Sandipan Dutta & TT (IBS)

Albert Libchaber (Rockefeller)

Jean-Pierre Eckmann (Geneva)

Michael Mitchell & Stanislas Leibler (IAS & Rockefeller)

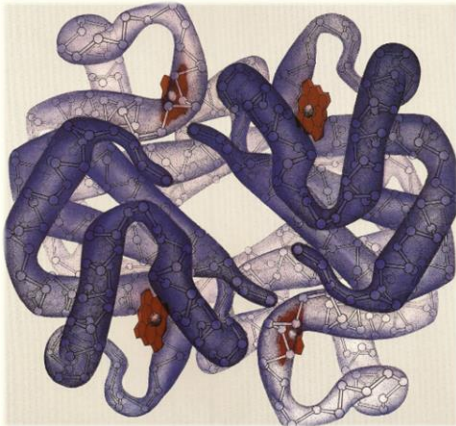
Proteins are complex systems made of hundreds to thousands amino acids arranged in a 3D structure



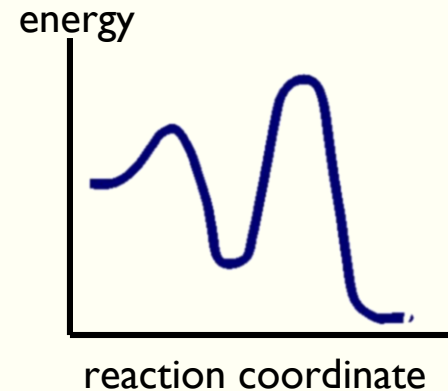
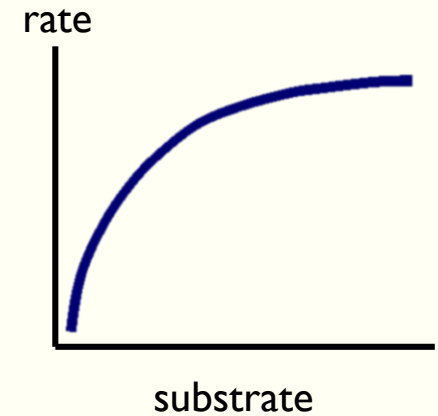
- » Non-random, information-rich.
- » Small, heterogeneous.
- » Collective interactions
- » Evolutionary memory.

Hemoglobin,
Irving Geiss, 1978

Genes and proteins have *many* degrees-of-freedom but protein function is described by a *few* parameters



**dimensional
reduction**



```

ATG CTG GAG AAC GGG CTG GCG CGG
Met Leu Glu Asn Gly Leu Ala Arg

TGG GAG CGG TTC CGC TGC AAC GTG
Trp Glu Arg Phe Arg Cys Asn Val

TGC ATC AGT GAG ATG CTC TTC ATG
Cys Ile Ser Glu Met Leu Phe Met

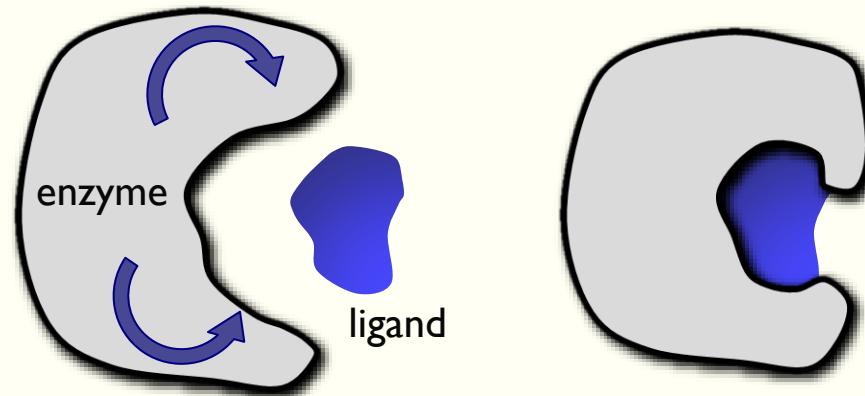
GAC GGC TGG AGG GAG CTG GGC TAC
Asp Gly Trp Arg Glu Leu Gly Tyr

TGG GCC GCC AAG CAG CGT GAC ACT
Trp Ala Ala Lys Gln Arg Asp Thr
    
```

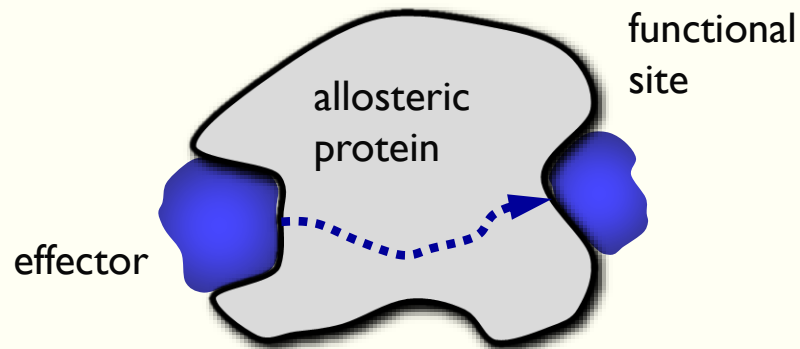
Information ~ gene length \longrightarrow Information ~ a few bits

Many proteins use *large-scale collective motions* to function

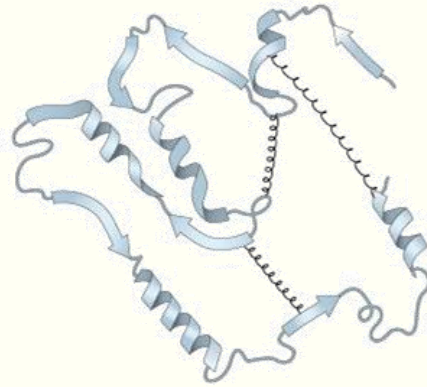
Induced fit:



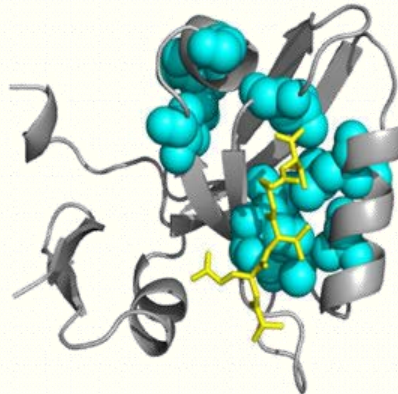
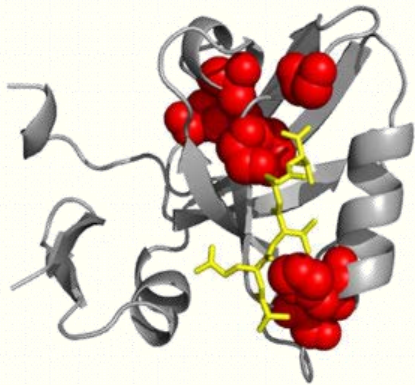
Allostery:



Protein sequences are also remarkably collective, with correlations indicating structure and function



Correlations predict 3D contacts.
Sander Marks Colwell Hopf et al.



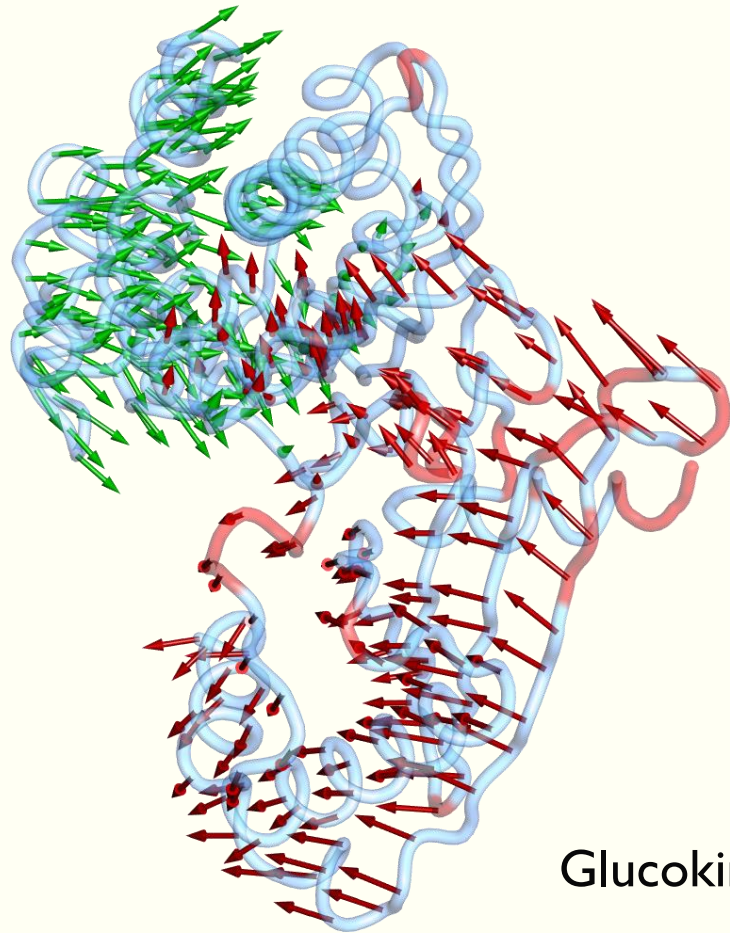
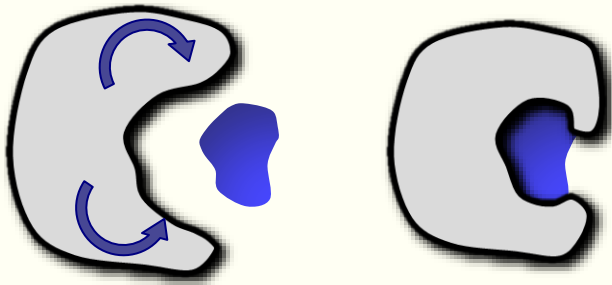
Sectors: coevolving protein domains
Halabi Rivoire Leibler Ranganathan et al.

Epistasis = interaction among mutations

Looking at dynamics suggests a mechanical basis for mapping genetic information to collective function

Motion (displacement):

unbound → bound

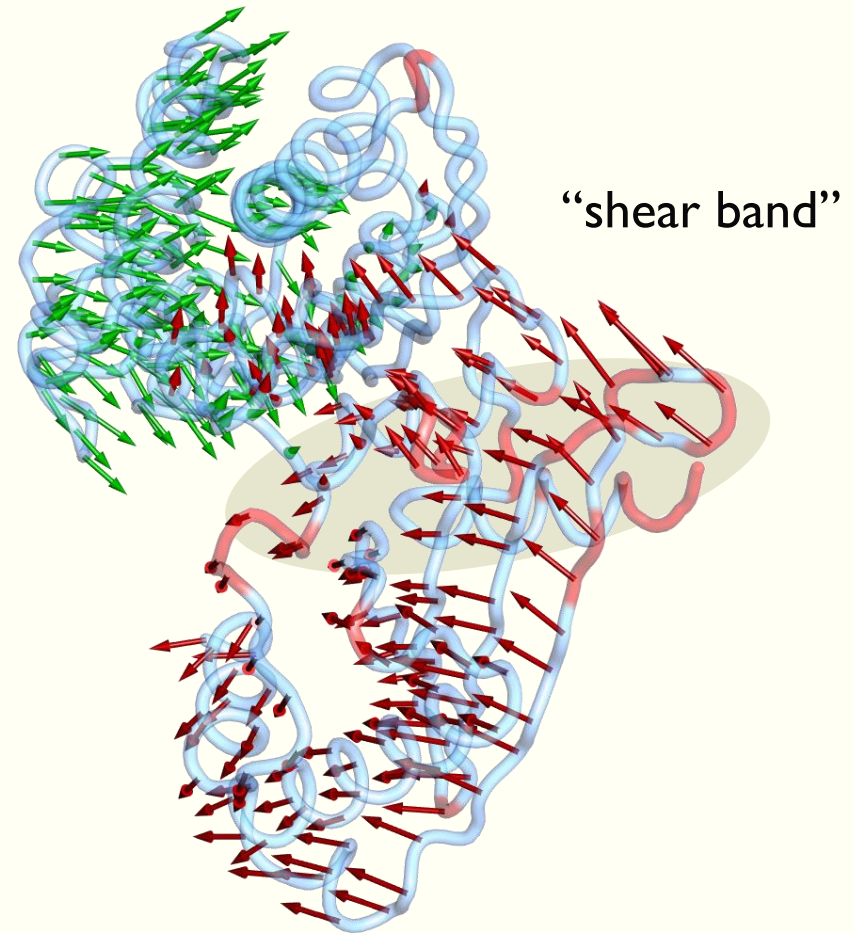


Glucokinase

The shear highlights the relative motion of amino acids within the protein

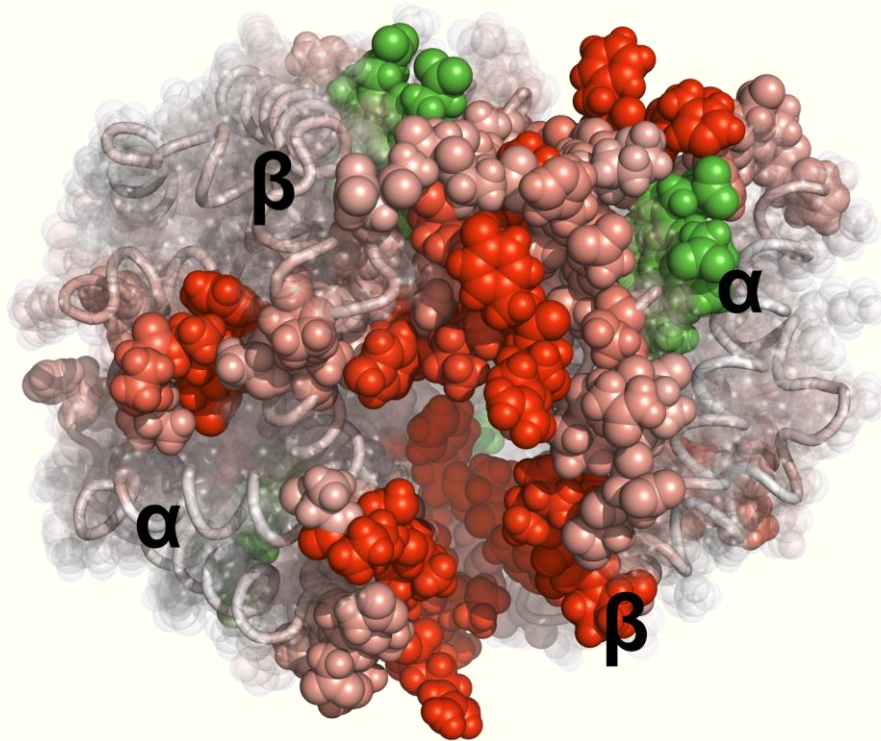
Shear = gradient (motion)

$$\epsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \sum_k \frac{\partial u_k}{\partial x_i} \frac{\partial u_k}{\partial x_j} \right)$$



Glucokinase (PDB: 1v4s → 1v4t)

A two-dimensional “shear band” communicates mechanical signals between distant sites

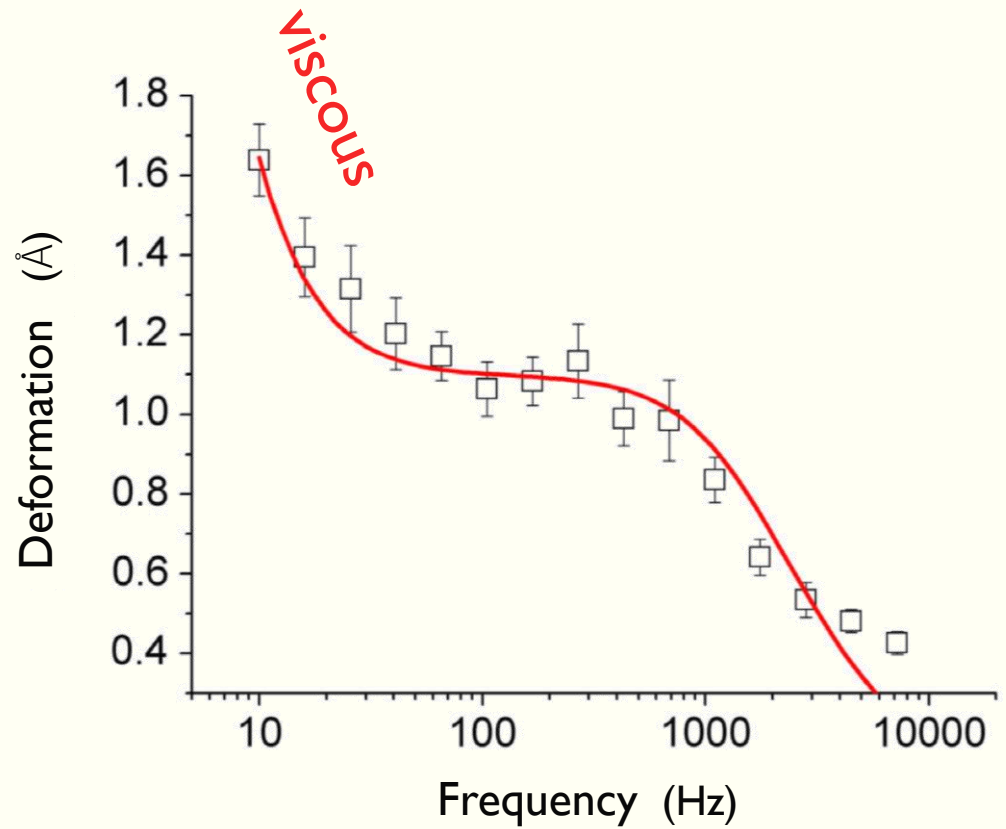
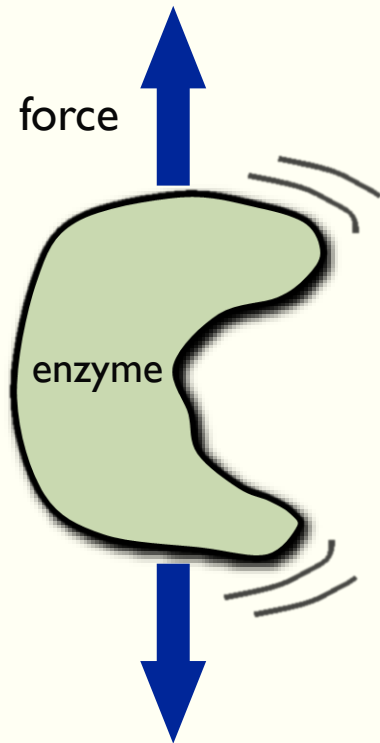


Hemoglobin

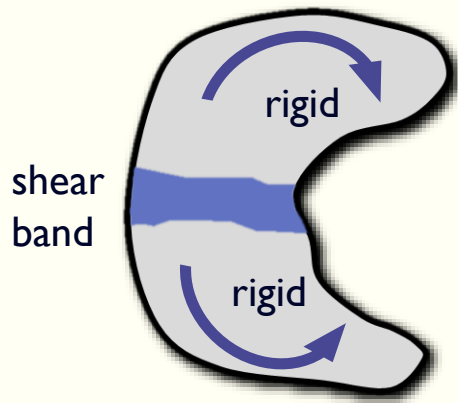
Shear band dimension

- Hemoglobin: 2.3
- Glucokinase: 2.0
- ATCase : 2.3
- Albumin : 1.9

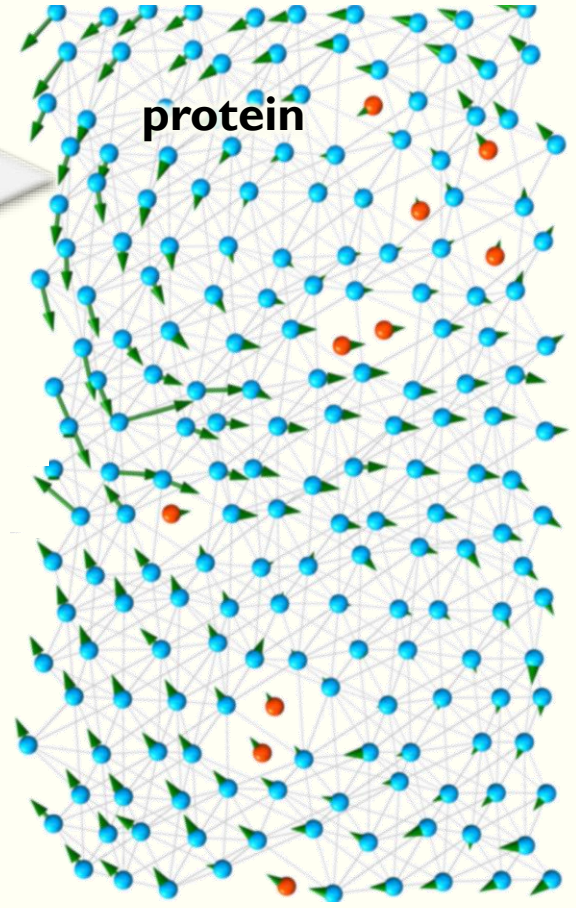
Proteins are soft and *viscous* at low frequency



Minimal model of protein as evolving matter to give evolution a mechanical interpretation



gene
 ...101111101111101...
 ...001111111001100...



| amino acids | H | P |
|-------------|--------|------|
| H | strong | weak |
| P | weak | weak |

aa = 200, z = 12

Coarse-grained network models of protein examine large-scale dynamics

Floppy modes in amorphous matter

» **“fracton” spectrum in glasses**

Alexander & Orbach. J de Phys (1982).

» **Floppy modes in amorphous solids**

Alexander. Phys Rep (1998).

» **Constraint theory in glasses**

Thorpe & Philips. Solid State Comm, (1985)

Elastic networks and normal modes in proteins

» **Protein normal-modes.**

Levitt, Sander & Stern. J Mol Bio (1985).

» **Large amplitude elastic modes in proteins.**

Tirion. PRL (1996).

» **Elastic network models**

Bahar et al. (many works since 1990s).

Allostery in elastic network models

» **Allostery in coarse-grained models of proteins.**

Min & Wall. PRL (2005).

» **Low-frequency modes describe allostery.**

Zheng, Brooks & Thirumalai. PNAS (2006).

» **Large-scale allosteric deformations**

Arora & Brooks. PNAS (2007).

Weakly-connected regions and viscoelasticity in proteins and biologically-inspired allosteric matter

» **“Protein-quakes”.**

Miyashita, Onuchic & Wolynes. PNAS (2003).

» **Viscoelastic properties of enzymes.**

Qu & Zocchi. PRX (2013).

» **Mechanical stress in functional enzymes.**

Joseph, Tseng, Zocchi & TT. PLoS I (2014).

» **Strain analysis of protein structures and low-dim.**

Mitchell, TT & Leibler. PNAS (2016)

» **Allostery-inspired response in mechanical networks.**

Rocks, Pashine, Bischofberger, Goodrich, Liu & Nagel. PNAS (2017)

» **Architecture and coevolution of allosteric materials.**

Yan, Ravasio, Brito & Wyart. PNAS (2017)

» **Elastic networks models of allosteric proteins.**

Flechsig. Biophys. J (2017)

» **Optimal cooperativity in allosteric materials**

Yan, Ravasio, Brito & Wyart. arxiv (2017)

Mechanical genotype-to-phenotype map of proteins

» **Self-referring DNA and protein.**

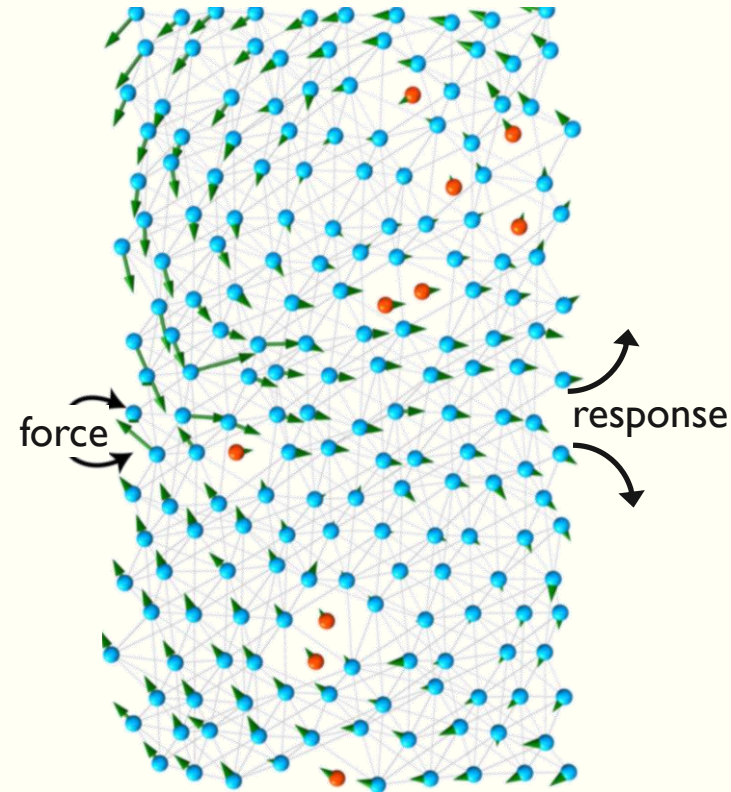
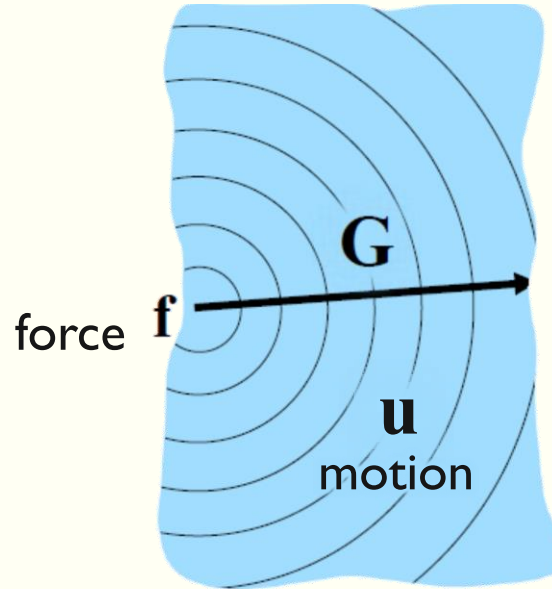
TT, Phil. Trans A (2016).

» **Physical model of genotype-to-phenotype map.**

TT, Libchaber & Eckmann. PRX (2017).

The Green function of the protein measures the transmission of force induced by a local impulse, such as ligand binding

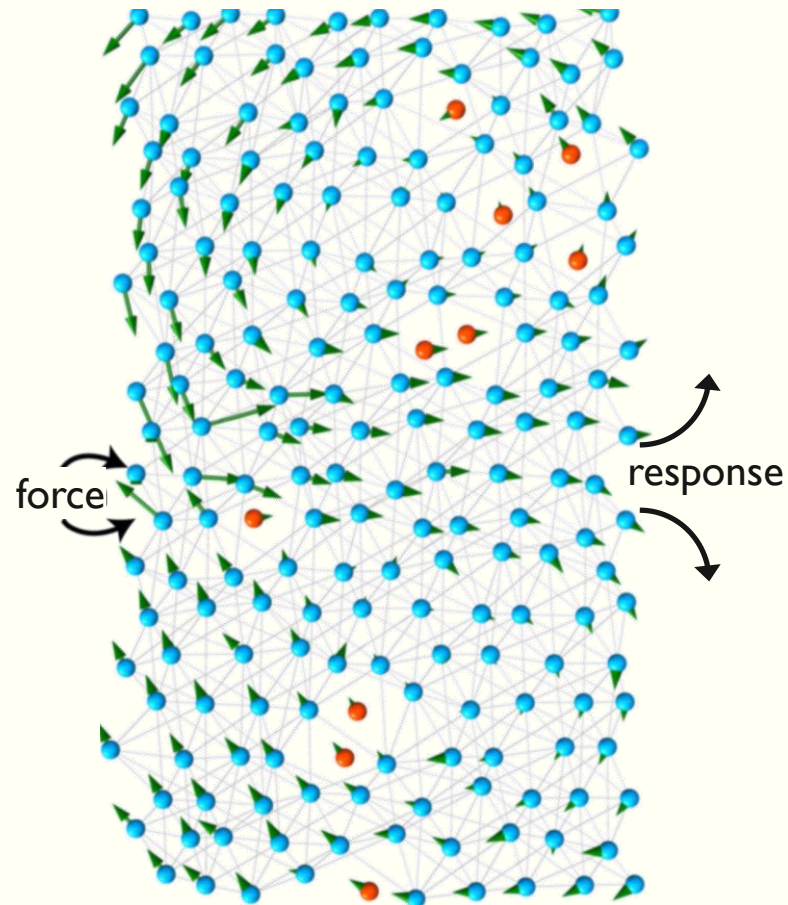
$$\mathbf{G}(\text{gene}) = \mathbf{G}(\text{aa net})$$



- Green's function solves the inverse problem

$$\left. \begin{array}{l} |\text{force}\rangle = \mathbf{H} |\text{motion}\rangle \\ |\text{motion}\rangle = \mathbf{G} |\text{force}\rangle \end{array} \right\} \Rightarrow \mathbf{G} = \mathbf{H}^{-1} \quad (\text{or } \mathbf{H}^+)$$

The Green function defines the genotype-to-phenotype map and the fitness landscape



Genotype-to-phenotype (gene \rightarrow motion)

$$|\text{motion}\rangle = \mathbf{G}(\text{gene})|\text{force}\rangle$$

Fitness landscape (gene \rightarrow fitness)

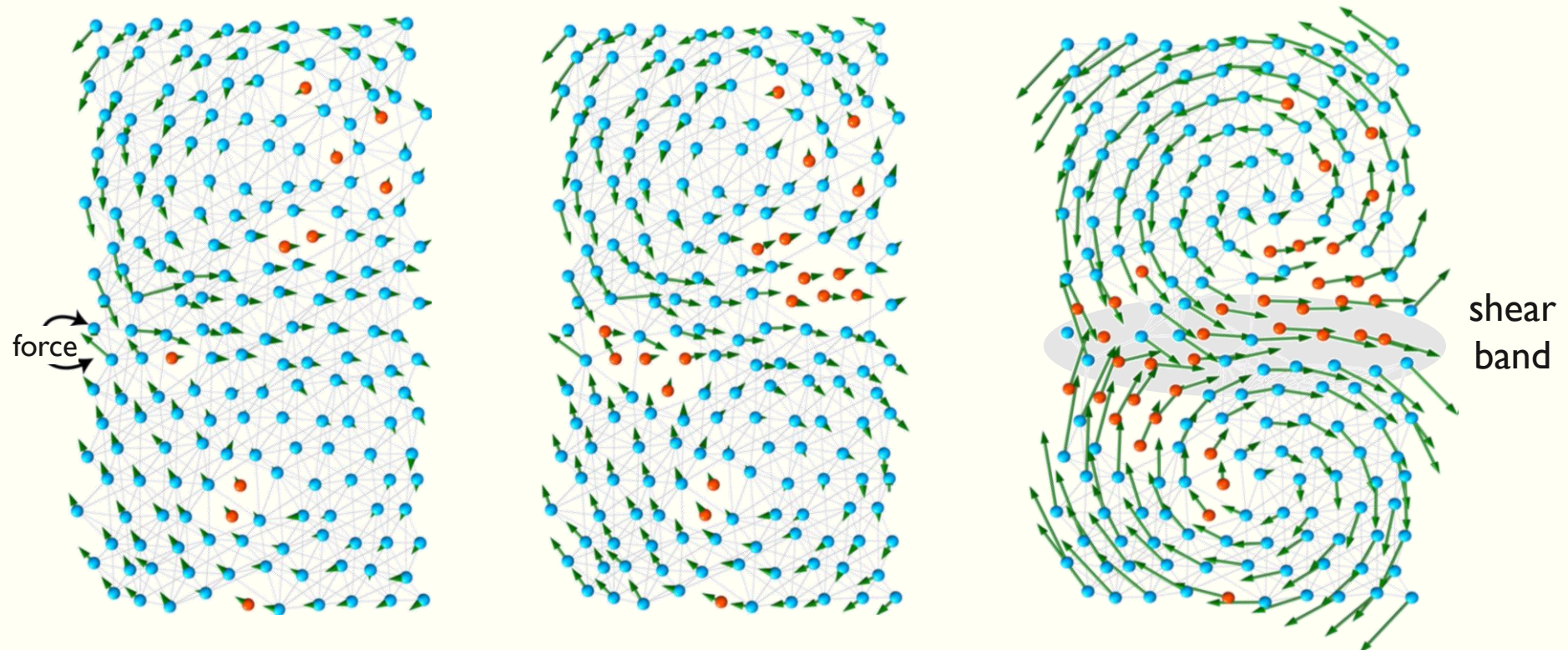
$$\begin{aligned} \text{Fitness}(\text{gene}) &= \langle \text{response} | \text{motion} \rangle \\ &= \langle \text{response} | \mathbf{G}(\text{gene}) | \text{force} \rangle \end{aligned}$$

Evolution searches in the mechanical fitness landscape for a shear band and soft mode

initial

intermediate

solution

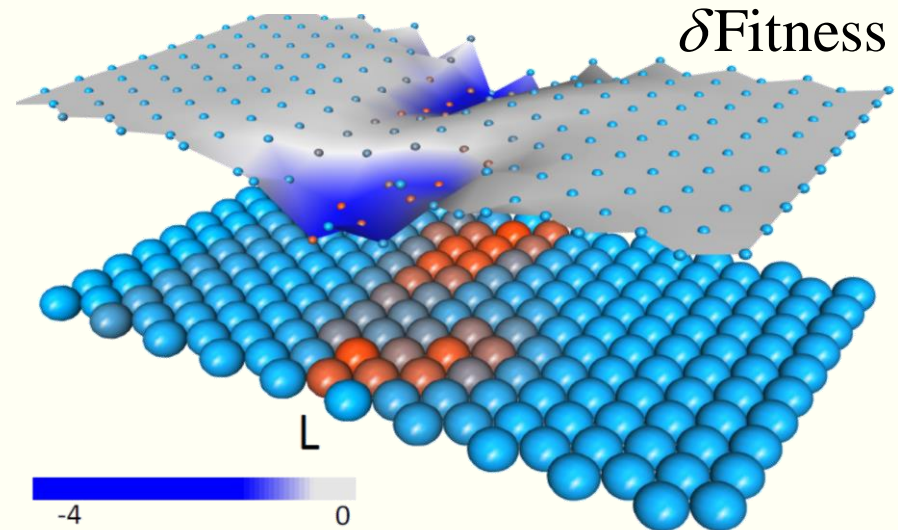
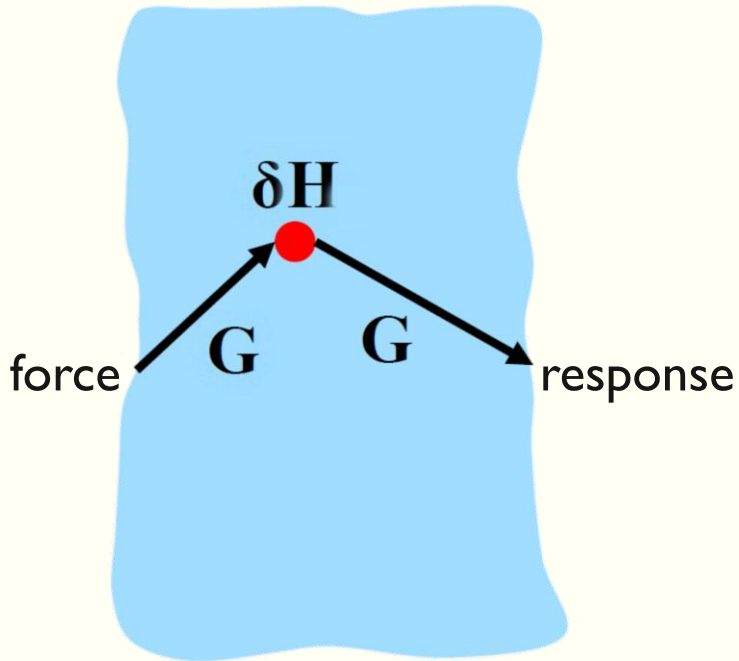


————— Fitness = $\langle \text{response} | \mathbf{G}(\text{gene}) | \text{force} \rangle$ —————>

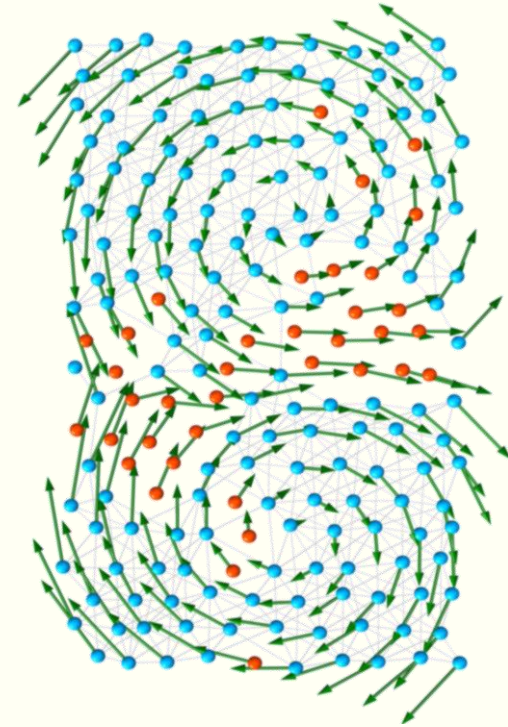
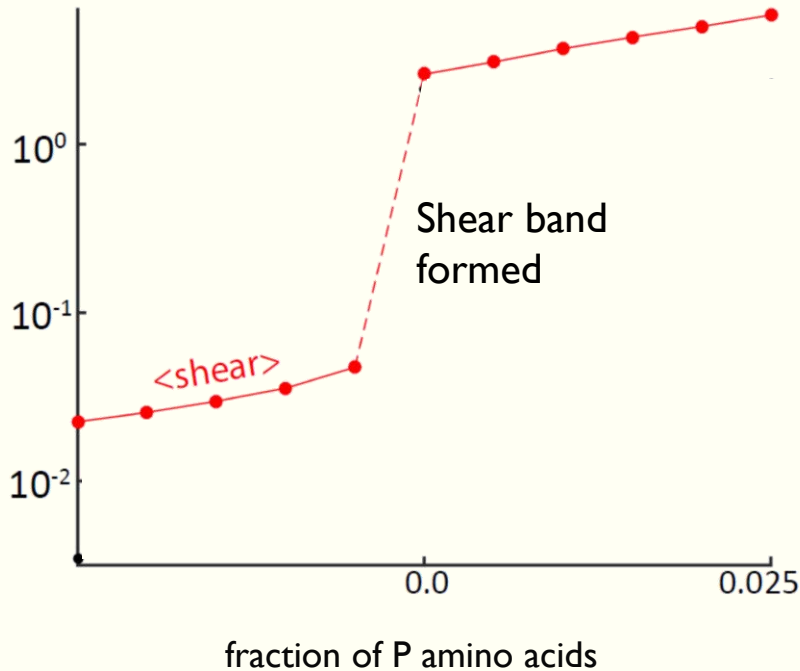
Mutations locally perturb the amino acid network, scatter the force, and thereby change the fitness

$$\delta G = -\mathbf{G} \delta H \mathbf{G} + \mathbf{G} \delta H \mathbf{G} \delta H \mathbf{G} - \dots$$

$$\delta \text{Fitness} = \langle \text{response} | \delta \mathbf{G} | \text{force} \rangle$$



Protein function emerges at a *topological transition* when a shear band of mutations divides the protein and a *soft mode* appears



$$\text{Fitness} = \langle \text{response} | \mathbf{G}(\text{gene}) | \text{force} \rangle \sim \frac{1}{\text{energy}(\text{soft mode})}$$

The genetic correlation among solutions increases towards the emergence of mechanical function

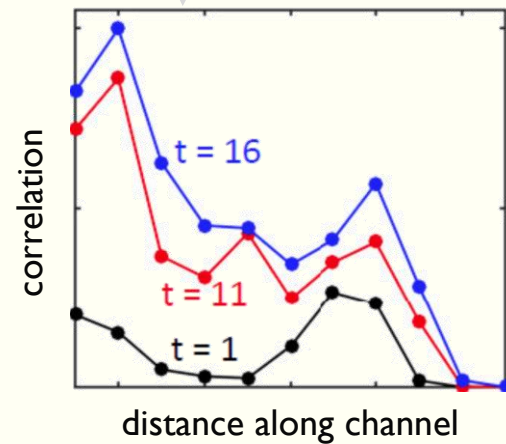
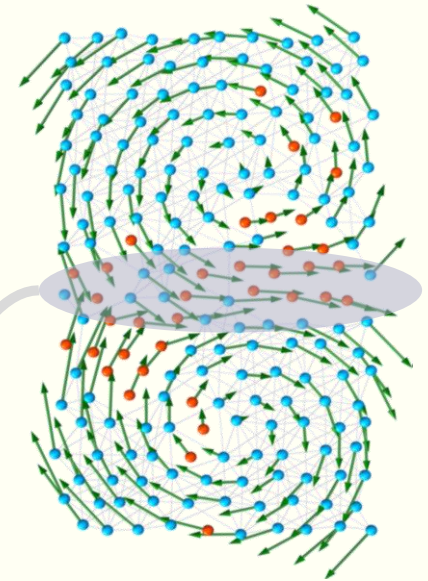
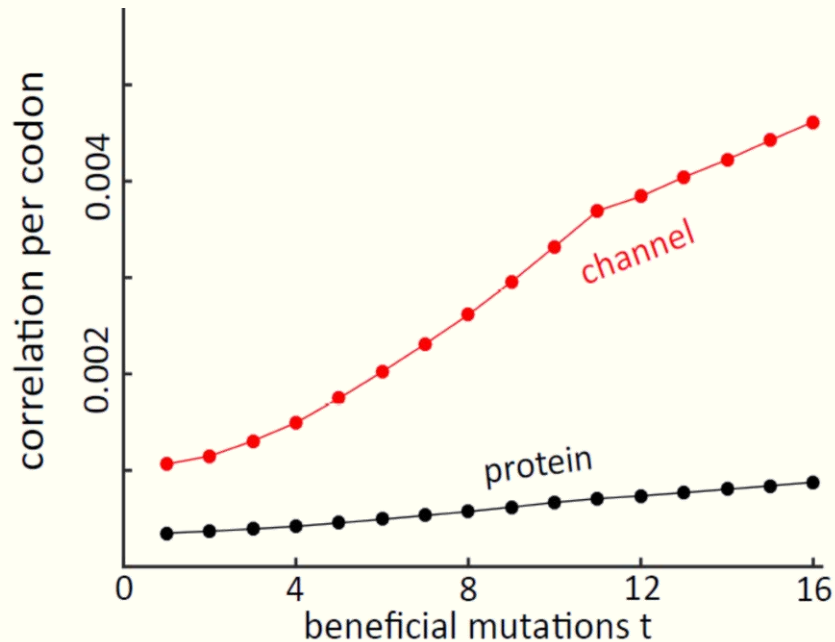
10^6 solutions

```

{
  11100101011111000010101...
  10010001000101011101110...
  00111110111110100111010...
  10011111110011001111000...
  ...

```

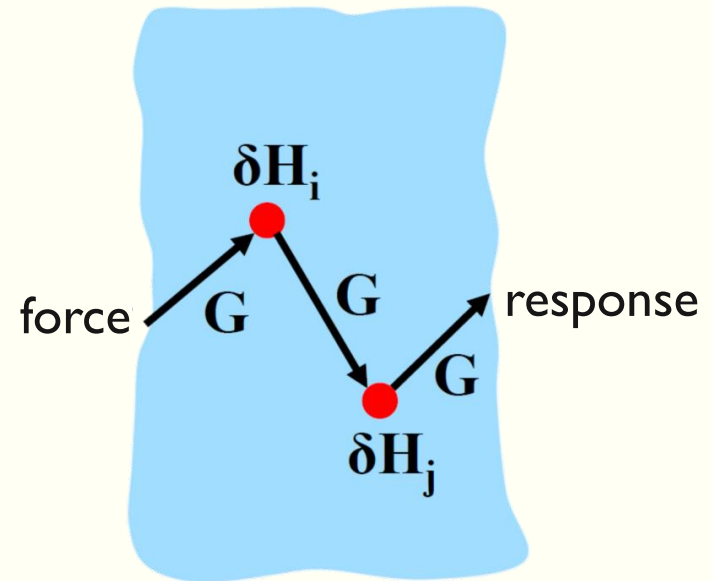
$$\text{correlation}_{ij} = \langle \text{gene}_i \text{gene}_j \rangle - \langle \text{gene}_i \rangle \langle \text{gene}_j \rangle$$



Epistasis is interaction of mutations resulting in nonlinearity of the mechanical fitness

$$\text{epistasis}_{i,j} = \delta\text{Fitness}_{i,j} - (\delta\text{Fitness}_i + \delta\text{Fitness}_j)$$

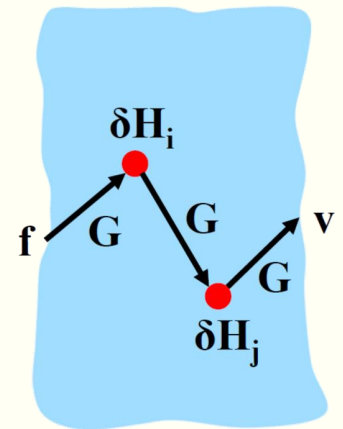
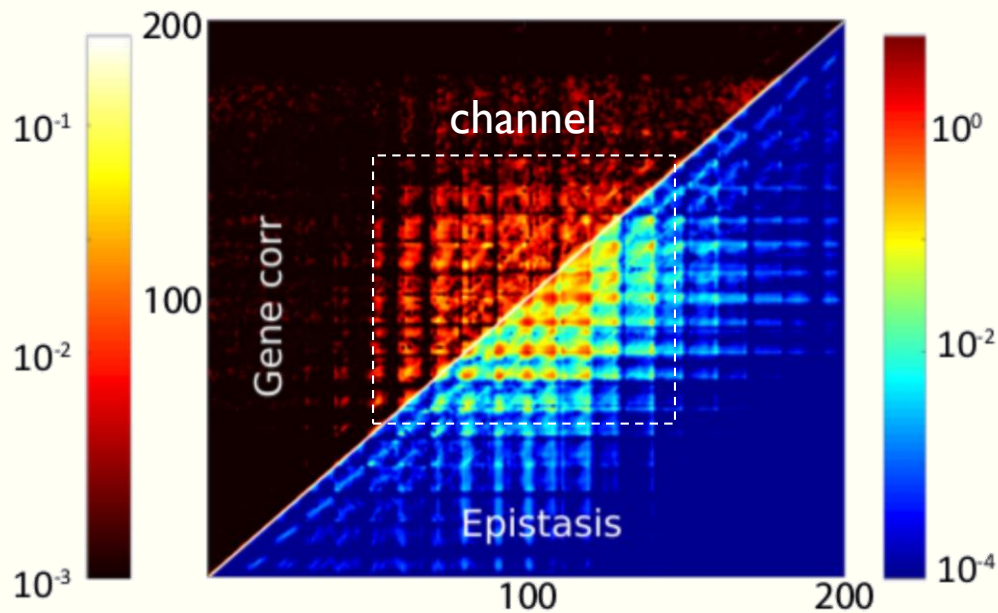
- Epistasis ~ “curvature” of G
 ~ sum over scattering paths:



$$\text{epistasis}_{i,j} = \langle \text{force} | \mathbf{G}\delta\mathbf{H}_i\mathbf{G}\delta\mathbf{H}_j\mathbf{G} + \mathbf{G}\delta\mathbf{H}_j\mathbf{G}\delta\mathbf{H}_i\mathbf{G}\dots | \text{response} \rangle$$

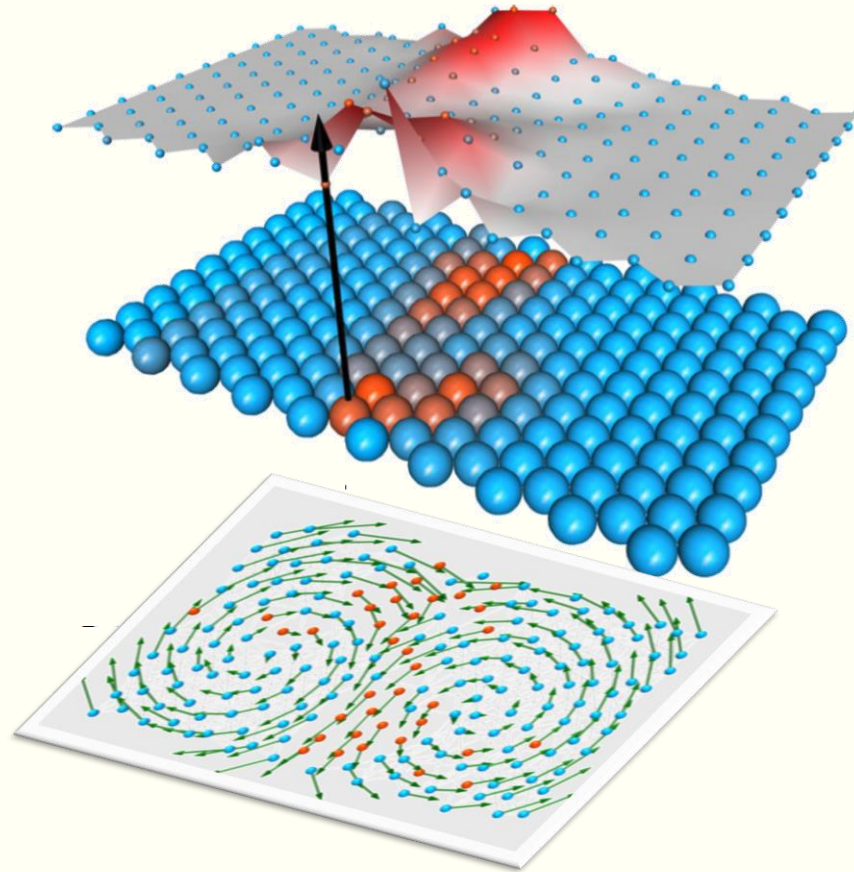
Mechanical forces link epistasis to genetic correlation, and both exhibit similar patterns

$$\text{correlation}_{ij} = \langle \text{gene}_i \text{gene}_j \rangle - \langle \text{gene}_i \rangle \langle \text{gene}_j \rangle$$



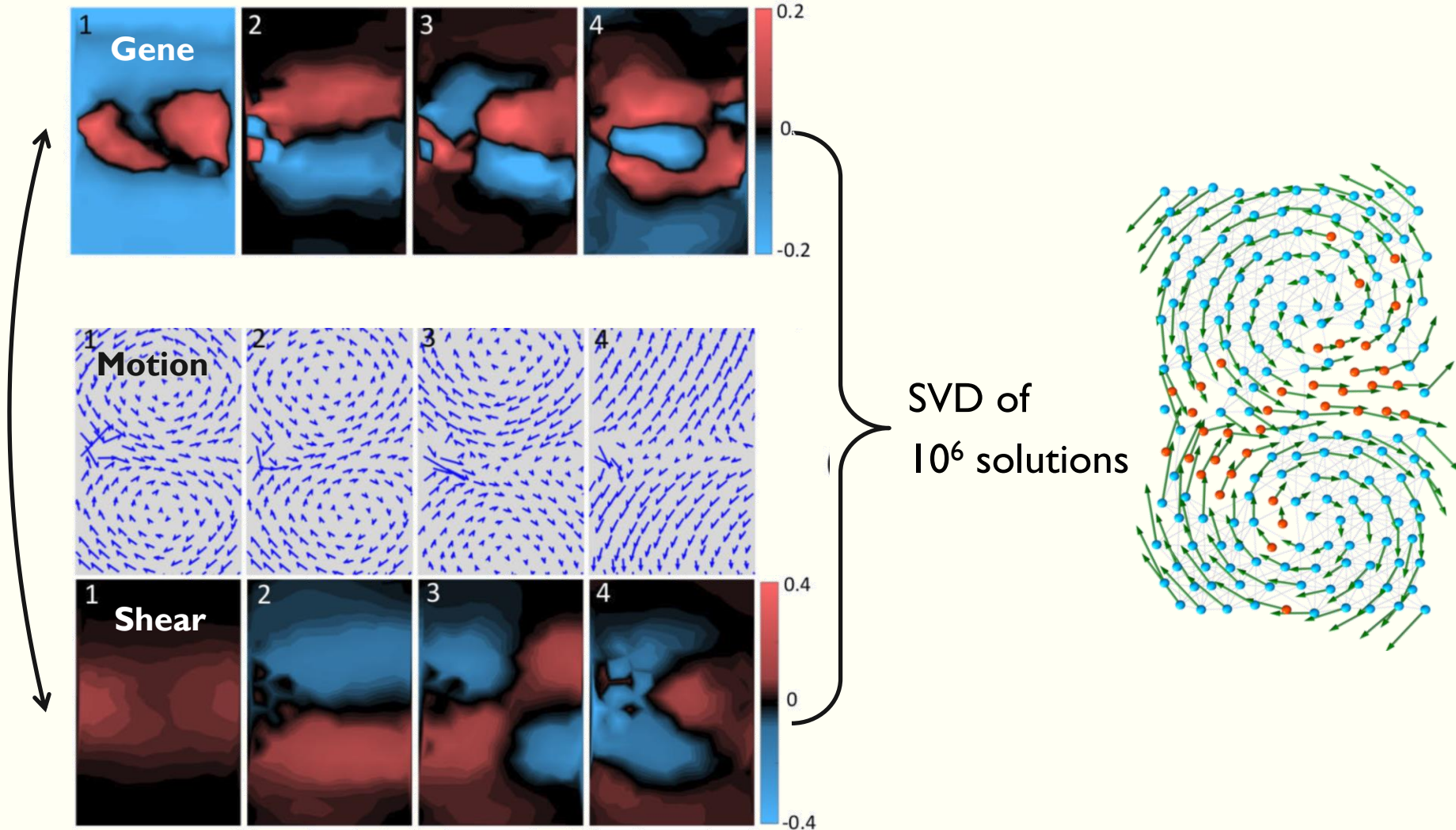
$$\text{epistasis}_{i,j} = \delta \text{Fitness}_{i,j} - \delta \text{Fitness}_i - \delta \text{Fitness}_j$$

Epistasis and sequence correlation are strongest in the high shear region separating the rigid subdomains



$$\text{epistasis}_{i,j} = \langle \text{force} | \mathbf{G}\delta\mathbf{H}_i\mathbf{G}\delta\mathbf{H}_j\mathbf{G} + \mathbf{G}\delta\mathbf{H}_j\mathbf{G}\delta\mathbf{H}_i\mathbf{G}\dots | \text{response} \rangle$$

Phenotype and genotype modes capture the relevant degrees-of-freedom of mechanical function and evolution

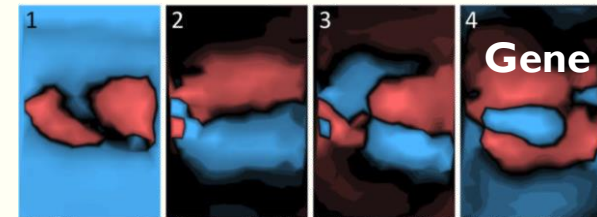
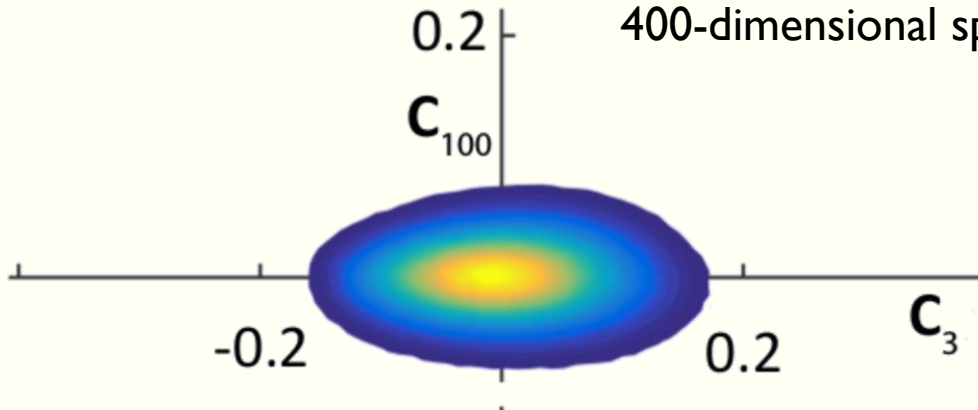


The genotype-to-phenotype map involves huge dimensional reduction

10^6 solutions

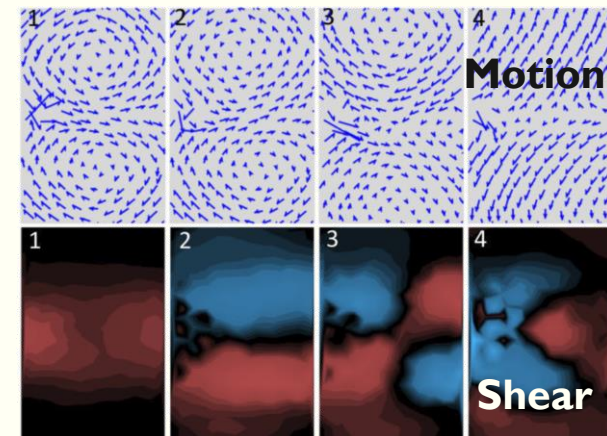
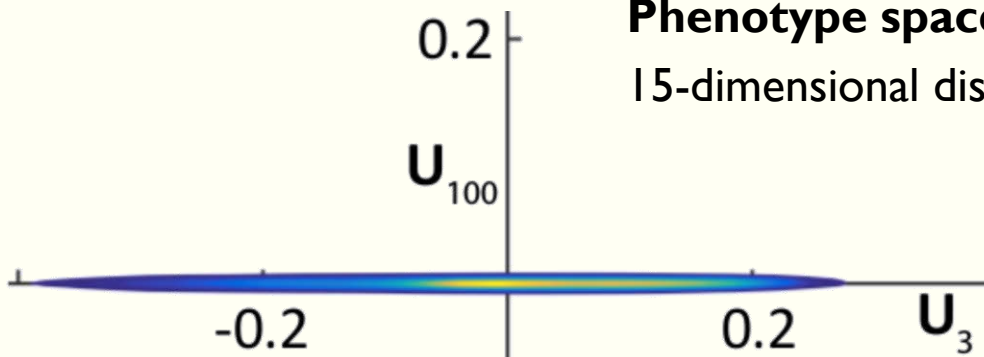
Genotype space

400-dimensional spheroid

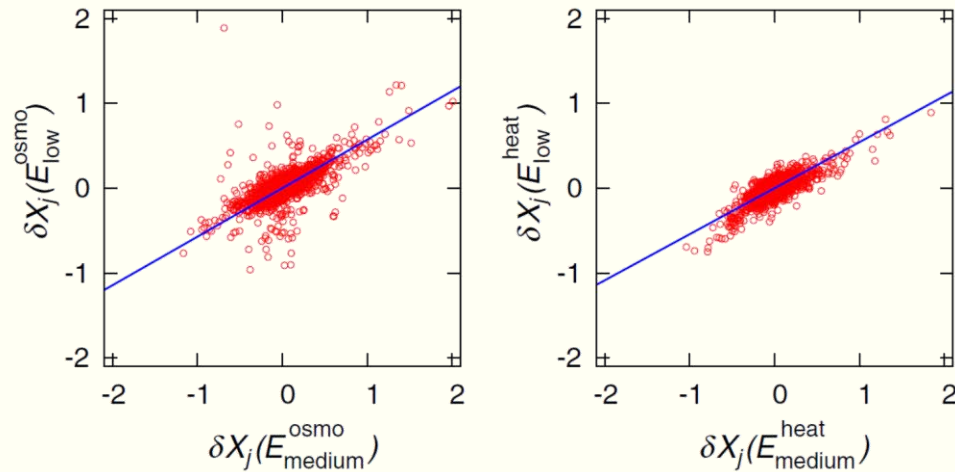


Phenotype space

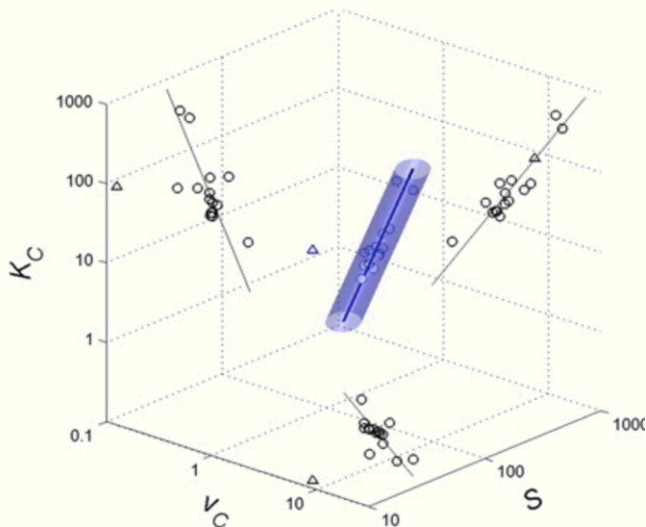
15-dimensional discoid



Dimensional reduction is a general phenomenon in life



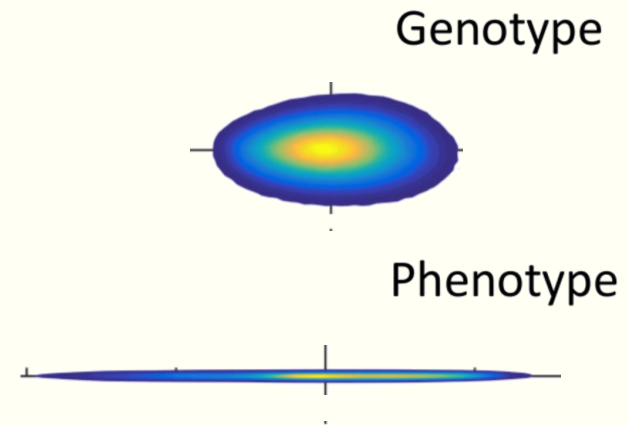
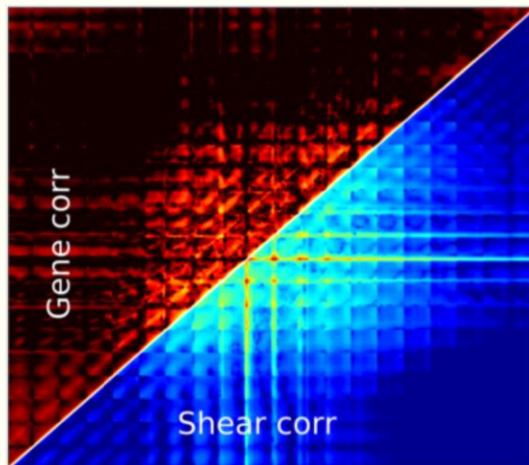
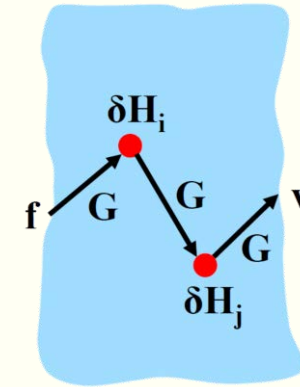
1D response of gene expression to stress
(Kaneko, Furusawa & Yomo, PRX 2015)



1D scaling of kinetic parameters in Rubisco
(Savir, Noor, Milo & TT, PNAS 2010)

Genetics and mechanical function are linked by a minimal theory

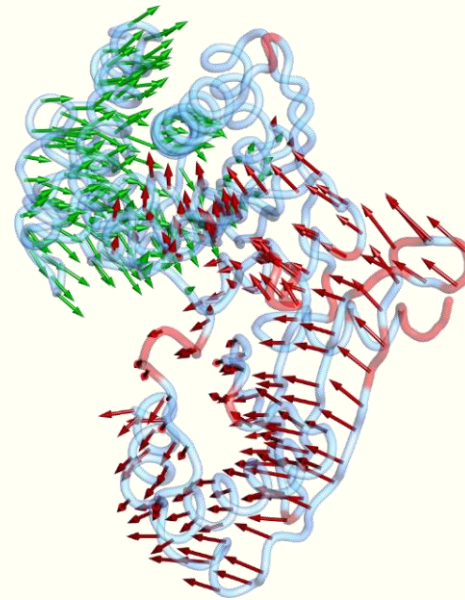
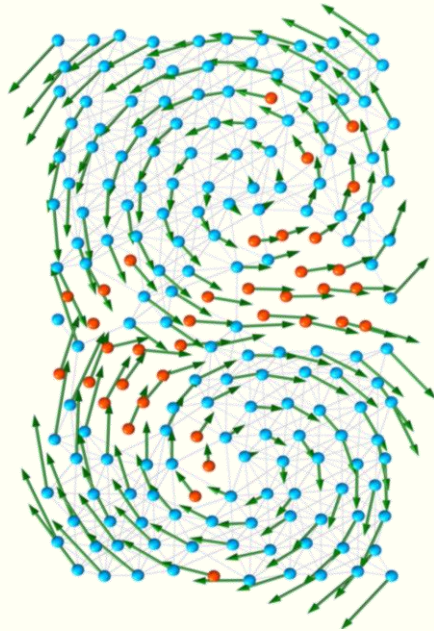
- **Mutations:** mechanical perturbations
- **Epistasis:** multiple scattering
- **Genotype-to-phenotype map**
- **Fitness landscape**
- **Soft modes at topological transition.**
- **Dimensional reduction**
- **Correspondence:** mechanical modes – coevolution modes



Dutta Eckmann Libchaber & TT, in review
TT Eckmann & Libchaber, Phys Rev X 2017

- **Applicable to other strongly-coupled living systems?**

Looking at protein dynamics suggests simple mechanical understanding of protein as evolving amorphous matter



Thanks:

S. Dutta (IBS)

J.-P. Eckmann (Geneva)

A. Libchaber (Rockefeller)

S. Leibler & M. Mitchell (IAS)



Center for Soft
& Living matter